In unit six, we're going to discuss the publication process as,well as some ethical issues in scientific publication.,In this 1st module, I'm going to talk about the problem of plagiarism.,I'll start with an example. There have been a number of high profile plagiarism scandals in politics.,E.g. in 2016, melania Trump gave a speech at the Republican National Convention that directly plagiarized a 2008 speech of Michelle Obamas.,I'm showing you here some of the texts from Melania Trump.,You can see that it is a direct copy of Michelle Obama’s words.,She said, obama said in 2008, you work hard for what you want in life, that your word is your bond, ,and you do what you say you're going to do, that you treat people with dignity and respect.,Trump in 2016 says, from a young age, ,my paren'ts impressed on me the values that you work hard for what you want in life,, that your word is your bond and you do what you say and keep your promise that you treat people with respect.,The chances that two people independently came up with these same words from scratch are infiniteismally small.,Trump has cut in paste obama's words and just made a few minor changes.,This is plagiarism. Plagiarism is when you try to pass off another person's writing as your own.,And I think most people are aware that you shouldn't do this, that it's unethical to plagiarize where it gets murky.,However is, I think many students aren't completely sure exactly what constitutes plagiarism.,In this module, I'm going to try to make that clear.,Most people know at a high level that you shouldn't be stealing large chunks of text from others.,But plagiarism can be more subtle than that, cutting and pasting a sentence or even a part of a sentence from another source and putting that into your own work as if you've written it.,That's plagiarism, taking somebody else's work and cutting and pasting it into your document and then slightly rearranging the words or changing a word here or there.,That's plagiarism barring descriptions or definitions directly from Wikipedia. That's plagiarism. When you are writing, you need to put things completely into your own words.,Or if you are going to borrow someone else's words, you need to put that material in quotation marks inside the source.,Many years ago, one of the 1st classes I taught at Stanford on women's health, I assigned essays, and I was shocked when I got the 1st essay pack from students.,Many students had taken sentences, phrases, sometimes even whole paragraphs, from the assigned readings from the course, without quotation marks and without citations.,Of course, I recognized all the words since I had assigned those readings and was very familiar with them.,I realized that many students were unaware that taking bits and pieces of other people's words and stringing them together in your essay without proper citation and quotes is plagiarism.,So I think many students are a bit hazy on the concept.,Let me give you a concrete example.,This is a hypothetical example. I made it up, but it's representative of what I've seen students do.,This is a Wikipedia entry on Earnest Hemingway.,It says Ernest Miller Hemingway was an American author and journalist.,His economical and understated style had a strong influence on 20th century fiction, while his life of adventure and his public image influenced later generation.,Now imagine a student is writing about Earnest Hemingway.,So they go to Wikipedia, and they start cutting and pasting bits and pieces from the Wikipedia entry, ,and they come up with the following Ernest hemingway's thrifty and understated style strongly influenced 20th century fiction.,His audacious lifestyle and public image also influenced later generations.,This is plagiarism. Notice that when I made up this version, I substituted a few words.,I substituted thrifty for economical. I substituted audacious lifestyle for life and of adventure.,And I rearranged a few things and changed a few other things, like, had a strong influence to strongly influenced.,But these changes are just cosmetic.,These are not my words. These are not my thoughts. These are the words and thoughts of the original author.,I could have rewritten that text without having any understanding of Hemingway, or of what a thrifty or understanded style even means, or what his lifestyle was.,In order to write something original about Hemingway, I can't just go to Wikipedia.,I have to go to enough sources so that I can figure out for myself about his style and influence.,I have to understand Hemingway enough to be able to generate new prose about him.,And that's what you need to do to avoid plagiarism.,When you're writing about others ideas or work, or you're writing from sources, you need to understand and digest the material well enough so that you can put it in your own words.,This means you're probably going to need to go to multiple sources to adequately understand,the subject matter. As you're collecting source material, it's okay to cut and pace passages from other people's work to get everything in one place, like in one,word document. but put those passages in quotation marks so you don't later forget that those aren't your words.,And when you go to write about it, work from memory.,Never start with someone else's text and just move their words around.,By the time you are ready to write, you should understand the material well enough to be able to start from scratch, draw your own conclusions, come up with your own ideas on the topic.,Don't trust that the person on Wikipedia has got it right.,And don't just mimic the original author sentence structure or just rearrange the original author's words.,That's still plagiarism. It's easy to detect plagiarism with Google.,If you ever suspect plagiarism, take a full sentence from the suspect material and put it into quotes, and then put that into Google.,If Google returns to you that same sentence in another paper, paper, it's almost certainly plagiarized.,The statistical probability of two people independently coming up with the identical, or near identical string of seven to ten words is astronomically low. There are common phrases in cliches that maybe four or five words long that are widely used.,But by the time you get to about seven words in a row, ,and certainly by the time you get to ten, the chances of two people independently stringing together those same ten words in a row are infinitely small.,In one case, I was editing a review article that a colleague could ask me to look at.,She had commissioned it from an expert, and she wanted someone to edit it.,As I was reading it, I got suspicious about plagiarism for various reasons, and I started doing some sluthing.,I took random sentences from the paper and put them in Google in quotes, and sure enough, I popped up all sorts of plagiarism.,The author had pulled one sentence from this paper, and two sentences from this other paper, and another few sentences from a 3rd paper.,The article was completely plagiarized, and once I started looking for it, it wasn't hard to sloth out that plagiarism.,I'll share with you another example of plagiarism that I happen to come across.,A while back, I was writing a paper on the use of estrogen to treat young women with bone loss due to eating disorders.,There were only a few previous randomized trials looking at this question, and I had all the papers in front of me on this subject, and I was reading this 1995 paper that you see here.,They had found a negative result.,Estrogen in their study did not help the bones of women with anorexia neurosa.,And as I was reading through the discussion, ,I read this paragraph that you see here, in which the authors laid out possible explanations for the negative results, and they listed the explanations with numbers.,They said, one possibility is this?,The 2nd possibility? Is this the 3rd possibility?,Is this? And a 4th and likely explanation is this?,So that structure actually stuck in my head.,A few days later, I was reading through a 2002 paper, and I got a sense of deja vu.,The discussion section of that later paper had the same numbered list of explanations.,So I rifled back through my pile of papers and found the 1995 paper, and started comparing the two papers. Here's what I found. The 2002 paper was by a different set of authors, and it was published in a different journal, ,but it was a similar study, another randomized trial, putting anorexic women on estrogen, and they also had a negative result.,I put the two papers by side by side, and I'm showing you here in red, ,the only differences between the 1995 paper and the 2002 paper in that particular passage, basically they were identical.,The 2002 authors just changed a few words here and there.,Instead of dose of estrogen, they said estrogen dose.,Instead of bone mass can they said, bone mass may.,Obviously those changes are not sufficient.,This is still blatant plagiarism. One funny thing is, here they, I think we're trying to cover their tracks a little.,So they took out the 3rd explanation the 1995 paperhead four explanations.,In the 2002 paper, they cut it down to three, but it's still plagiarism ully comparing their contents.,The 2002 paper was almost entirely a direct copy of the 1995 paper.,The data were different, but almost all of the text of the 2002 paper was copied straight from the 1995 paper.,Just to show you a few examples here, I'm showing you the concluding sentences of both,papers. You can compare them on your own, but you can see that they are nearly identical.,Just one more example from this same case, if you want to look at it again.,Another example of just the blatant plagiarism in that 2002 paper to be fear that 2002 paper was coming from,a foreign country, and my guess is that the authors weren't confident in their writing of English, so they decided to take that 1995 paper and sort of use it as a template.,Of course, that doesn't excuse the behavior, but that might explain what happened.,This is a very extreme case of plagiarism, but it's actually published in the literature.,Hopefully, modern plagiarism detection software will prevent such extreme cases in the future.,Another thing that students are often less aware of is the concept of self plagiarism.,This is recycling your own writing or your own data from one published paper to the next. That constitutes self plagiarism, and it's also unethical.,The problem is that if you're just rehashing old material, why are you even publishing a new paper at all?,If you have to rehash old material, that means you don't have anything new to say.,Plus you may be violating copyright laws from the journal that owns the published paper.,You are not supposed to plagiarize, even from your own work.,Now, one possible exception to this is that there may be some duplication of text within the materials and method section, E.G.,There may be only so many ways to explain what you did to the cells.,So if you've used the same experiment across multiple studies, there may be some text repeated.,It shouldn't be a complete copy.,But if you explain some of the materials and methods in the same way in multiple papers, journal editors are usually fine with that, but everything else should be new.,The introduction section shouldn't just be a rehash of a previously published introduction.,E.g., authors also sometimes engage in something that's akin to plagiarism, which is duplication of their old results in their old data.,Sometimes an author will take a dataset that's already published and maybe add a little bit of new data, or tweak the data slightly and present the whole thing as a new paper.,This is misleading, because its putting things out there as if there were two independent pieces of information supporting a hypothesis when there isn't.,It's patting the medical literature with superfluous papers.,And there's also, of course, a copyright issue.,So self plagiarism also is not allowed.,People have actually tried to go through the literature and figure out exactly how prevalent is plagiarism in the scientific literature.,There was a story in Nature where they reported that in pilot tests of the plagiarism detection software cross check,, some journals found that 6% to 23% of submitted manuscripts had to be rejected outright due to plagiarism.,6% to 23% is a surprisingly high percentage, I think.,Now this includes both plagiarism and self plagiarism. Another study used automatic detection software, and then they manually confirmed plagiarism in submitted papers.,They found that 8% of the papers had plagiarized other people's work, and about 3% involve self plagiarism.,So again, we're arranging around 10% of the papers having significant plagiarism in there.,That's a pretty big number. Another group of researchers did a study of plagiarism in the personal statements for residency applications.,These were doctors who were applying to do their residency at Brigham and women's Hospital, which is a very prestigious institution.,The researchers used a plagiarism detection software to review about 5000 personal statements, and then they confirmed suspected plagiarism.,Manually, they found that 5% of essays had clear evidence of plagiarism, meaning one in 20 of the doctors applying for this prestigious residency had plagiarized.,So plagiarism is a significant problem.,Be careful in your own work, to put everything in your own words.,I want to share one last example that's a bit more subtle.,I suspect that this kind of plagiarism is more common than the more blatant examples of plagiarism that I shared earlier.,So it's worth me wrapping up with here.,I was reading a 2009 paper, and I came across this sentence.,Recent registry based and hospital based studies have documented a statistically significant increased risk of melanoma after breast cancer, ,with standardized incidents ratios ranging from 1.4 to 2.7.,I was writing about this topic for the lay public, and I wanted to verify the numbers so that I could use them in my story.,So I pulled some of the references from the 2009 paper.,When I pulled their 5th reference, a paper from 2004, I found the,exact same sentence, which I'm underlining here.,The 2009 authors had copied word for word,,and reference for reference one, two, three and four.,Those references are the exact same, uh, the summary of the 2004 authors.,Now, this is subtle, and to be fair, I didn't find any other evidence of plagiarism elsewhere in the paper. But this is plagiarism, and there are two problems with it.,1st, the 2009 authors have borrowed the work and exact words of the previous authors.,The previous authors went through the literature, summarized everything, synthesized, did it and put it into words, and the 2009 authors are just stealing that.,The 2nd problem is that the 2009 authors are completely trusting that the 2004 authors have accurately summarized the literature.,That's a leap of faith. I wouldn't take As I've told you before, MS.,citations and inaccurate citations are rampant in the literature.,I would never trust a secondary source like this.,I would always go back to the primary sources and check the numbers for myself.,Plus, the 2009 authors are assuming that nothing new has happened between 2004 and 2009, they haven't bothered to update the literature search.,Now, the authors of the 2nd paper could simply have put the copied sentence in quotation marks.,This acknowledges that they are borrowing the work and words of the previous authors, ,and also gets them off the hook with regards to accuracy, because by quoting, you are saying, ,these are the numbers according to someone else, you are not saying that you have checked or agree with those numbers.,But even better, what the 2009 authors should have done is to dig through the literature themselves and summarize the literature Denovo, rather than simply copying somebody else's summary.

在第六单元中，我们将讨论出版过程以及科学出版物中的一些伦理问题。在第一个模块中，我将讨论抄袭问题。我将从一个例子开始。政治上出现了许多备受瞩目的抄袭丑闻。例如，2016年，梅拉尼娅·特朗普在共和党全国代表大会上发表演讲，直接抄袭了米歇尔·奥巴马斯2008年的演讲。我在这里给你看梅拉尼娅·特朗普的一些文字。你可以看到这是米歇尔·奥巴马的话的直接副本。奥巴马在2008年说，你为生活中想要的东西而努力工作。你的诺言就是你的纽带，你要照你说的去做。你要以尊严和尊重的态度对待别人。特朗普在2016年说，从很小的时候起，我的父母就给我留下了深刻的印象，那就是你为生活中想要的东西而努力奋斗。你的诺言就是你的纽带，你照你说的去做，信守诺言。你以尊重的态度对待别人。两个人从头开始独立想出同样的词的可能性微乎其微。特朗普剪切并粘贴了奥巴马的话，只是做了一些细微的改动，这就是抄袭。抄袭是指你试图将他人的写作假冒为自己的写作。而且我想大多数人都知道你不应该这样做，抄袭是不道德的。但是，变得模糊的地方在于，我认为许多学生并不完全确定究竟是什么构成抄袭。在本模块中，我将尝试阐明这一点。大多数人都知道，在高层次上，你不应该从他人那里窃取大量文字，但是抄袭可能比这更微妙。剪切和粘贴一句话，甚至是来自其他来源的句子的一部分，然后像写作一样把它放到自己的作品中，那就是抄袭。拿走别人的作品，把它剪切并粘贴到你的文档中，然后稍微重新排列单词或者在这里或那里修改一个单词，那就是抄袭。直接从维基百科借用描述或定义，那就是抄袭。当你写作时，你需要把东西完全写成自己的话，或者如果你要借用别人的话，你需要将这些材料放在引号中并引用来源。很多年前，这是我在斯坦福大学教授的第一堂关于女性健康的课程。我分配了论文，当我从学生那里拿回第一篇文章时，我感到非常震惊。许多学生从课程中分配的读物中摘取了句子、短语，有时甚至是整段段落，没有引号，也没有引文。当然，自从我分配这些读数以来，我认出了所有的单词，并且对它们非常熟悉。我意识到许多学生没有意识到，如果没有适当的引文和引文，就把他人的话拼凑在你的论文中，是抄袭。所以我认为很多学生对这个概念有点模糊。让我举一个具体的例子。这是一个假设的例子。我编造的，但它代表了我所看到的学生的所作所为。这是维基百科上关于欧内斯特·海明威的条目。它说，欧内斯特·米勒·海明威是一位美国作家和记者。他经济而低调的风格对20世纪的小说产生了强烈的影响，而他的冒险生活和公众形象影响了后代。现在，想象一下一个学生在写关于欧内斯特·海明威的文章。于是他们去了维基百科，他们开始剪切和粘贴维基百科条目中的点点滴滴，他们想出了以下内容。欧内斯特·海明威节俭而低调的风格强烈影响了20世纪的小说。他大胆的生活方式和公众形象也影响了后代，这就是抄袭。请注意，当我编造这个版本时，我用了几个字代替。我用节俭代替了经济。我用大胆的生活方式代替了冒险生活，我重新安排了一些事情，改变了其他一些东西，比如产生了强烈的影响力，但这些变化只是表面上的。这些不是我的话。这些不是我的想法。这些是原作者的话语和想法。我本可以在对海明威、节俭或低调的风格到底意味着什么，或者他的生活方式一无所知的情况下重写那段文字。为了写一些关于海明威的原创文章，我不能只去维基百科。我必须找到足够的消息来源，这样我才能自己弄清楚他的风格和影响力。我必须足够了解海明威才能写出关于他的新散文，而这正是避免抄袭所需要做的。当你写关于他人的想法或作品的文章，或者你从来源写作时。你需要充分理解和消化这些材料，这样你才能用自己的话说出来。这意味着你可能需要访问多个来源才能充分理解主题。当你收集原始资料时，可以剪切和粘贴他人作品中的段落，将所有内容放在一个地方，就像在一个单词文档中一样。但是把这些段落放在引号里，这样你以后就不会忘记那些不是你的话。而且，当你去写这篇文章时，要记住记忆，永远不要从别人的文字开始，而只是四处移动他们的话。当你准备好写作时，你应该足够了解材料，以便能够从头开始。得@@出自己的结论，对这个话题提出自己的想法。不要相信维基百科上的那个人说得对。而且，不要只是模仿原作者的句子结构或只是重新排列原始作者的话。那仍然是抄袭。使用Goog@@le可以很容易地发现抄袭行为。如果你怀疑抄袭，请从可疑材料中取一整句话并加上引号，然后将其放入谷歌。如果谷歌在另一篇论文中向你返回同样的句子，那么几乎可以肯定，它被抄袭了。两个人独立得出相同或几乎相同的七到十个单词的字符串的统计概率低得惊人。有些常用的短语和陈词滥调可能长达四五个单词，它们被广泛使用。但是，当你连续写出大约七个单词时，当然当你达到十个单词时，两个人连续独立地将同样的十个单词串在一起的可能性微乎其微。有一次，我正在编辑一篇评论文章，一位同事让我看这篇文章。她委托专家撰写了这篇文章，她希望有人来编辑它。当我读这本书时，出于各种原因，我对抄袭产生了怀疑，于是我开始进行一些侦查。我从报纸上随机取了句子然后用引号把它们放在谷歌里，果然，我发现了各种各样的抄袭。作者从这篇论文中摘了一句话，从另一篇论文中摘了两句话，从第三篇论文中又摘了几句话。这篇文章被完全抄袭。而且，当我开始寻找它时，发现这种抄袭行为并不难。我将与大家分享我碰巧遇到的另一个抄袭的例子。不@@久前，我正在写一篇关于使用雌激素治疗因饮食失调而骨质流失的年轻女性的论文。之前只有几次随机试验研究这个问题，我面前有关于这个主题的所有论文。我正在阅读你在这里看到的这篇1995年的论文，他们发现了一个负面的结果。研究中的雌激素对神经性厌食症女性的骨骼没有帮助。当我通读讨论时，我读了这段话，你会在这里看到，作者在其中阐述了负面结果的可能解释。他们用数字列出了解释。他们说一种可能性是这个，第二种可能性是这个，第三种可能性是这个，第四种可能的解释是这样。所以那个结构实际上卡在了我的脑海里。几天后，我正在通读2002年的一篇论文，我有一种似曾相识的感觉。后来的那篇论文的讨论部分有同样编号的解释清单。于是我翻阅了一堆论文，找到了1995年的论文，然后开始比较这两篇论文。这是我发现的。2002年的论文由另一组作者撰写，并发表在不同的期刊上。但这是一项类似的研究，另一项让厌食症女性服用雌激素的随机试验，结果也为阴性。我把两篇论文并排放置，我在这里用红色向你展示那段特定段落中1995年的论文和2002年的论文之间的唯一区别。基本上它们是相同的。2002年的作者只是在这里和那里改了几个字。他们说的不是雌激素剂量，而是雌激素剂量。他们说骨量可以，而不是骨量可以。显然，这些变化是不够的。这仍然是公然的抄袭。这里有一件有趣的事情，我想他们想稍微掩盖自己的踪迹，所以他们拿出了第三种解释。1995年的论文有四种解释。在2002年的报纸中，他们将其减少到三个，但它仍然是抄袭。那时我很好奇，我仔细阅读了两篇论文，比较了它们的内容。2002年的论文几乎完全是1995年报纸的直接副本。数据有所不同，但2002年论文的几乎所有文本都是直接从1995年的论文中复制的。为了给大家举几个例子，我给你看两篇论文的最后一句话。你可以自己比较它们，但你可以看到它们几乎相同。再举一个来自同一个案例的例子。如果你想看的话，那又是2002年那篇论文中公然抄袭的又一个例子。公平地说，2002年的论文来自国外，我的猜测是作者对自己的英语写作没有信心。因此，他们决定把1995年的论文当作模板。当然，这并不能解释这种行为的借口，但这也许可以解释发生了什么。这是一个非常极端的抄袭案例，但它实际上已发表在文献中。希望现代抄袭检测软件能够防止将来出现这种极端情况。学生通常不太了解的另一件事是自我抄袭的概念。这是将你自己的写作或自己的数据从一篇已发表的论文回收到另一篇已发表的论文。这构成了自我抄袭，也是不道德的。问题在于，如果你只是在重温旧材料，为什么还要发表一篇新论文？如果你必须重温旧的材料，那就意味着你没什么好说的。另外，您可能违反了拥有已发表论文的期刊的版权法。即使是你自己的作品，你也不应该抄袭。现在，一个可能的例外是，“材料和方法”部分中可能有一些重复的文本。例如，可能只有这么多方法可以解释你对细胞的所作所为。因此，如果您在多项研究中使用了相同的实验，则可能会重复一些文本。它不应该是一份完整的副本，但是如果你用同样的方式解释一些材料和方法，在多篇论文中，期刊编辑通常可以接受。但是其他一切都应该是新的。例如，简介部分不应该只是对之前发布的简介的重述。作者有时还会从事类似于抄袭的事情，即重复他们的旧结果和旧数据。有时，作者会获取已经发布的数据集，然后可能添加一点新数据，或者稍微调整数据，然后将整个数据作为一篇新论文呈现。这具有误导性，因为它把事情当作两条独立的信息来支持假设，而事实并非如此。它在医学文献中充斥着多余的论文，当然还有版权问题。因此，自我抄袭也是不允许的。实际上，人们试图浏览文献，弄清楚抄袭在科学文献中到底有多普遍。《自然》杂志上有一篇报道，他们报道说，在抄袭检测软件CrossCheck的试点测试中，一些期刊发现，由于抄袭，提交的手稿中有6％至23％必须被拒绝。我认为，6％至23％是一个出人意料的高百分比。现在，这包括抄袭和自我抄袭。另一项研究使用了自动检测软件，然后他们手动确认了提交的论文中存在抄袭行为。他们发现，8％的论文抄袭了他人的作品，大约3％的论文涉及自我抄袭。因此，再说一遍，我们对大约10％的论文中存在严重抄袭行为进行了调查。这是一个相当大的数字。另一组研究人员对居留申请个人陈述中的抄袭行为进行了研究。这些医生正在申请在布里格姆妇女医院住院，这是一家非常负盛名的机构。研究人员使用抄袭检测软件审查了大约5000份个人陈述，然后他们手动确认了可疑的抄袭。他们发现，有5％的文章有明显的抄袭证据。这意味着在申请这项享有盛誉的住院医师中，每20名医生中就有1名是抄袭的。因此，抄袭是一个重大问题。在自己的工作中要小心，把所有内容都用自己的话说出来。我想分享最后一个更微妙的例子。我怀疑这种抄袭比我之前分享的更公然的抄袭例子更为普遍。因此，值得我在这里总结一下。我正在读一篇2009年的论文，我看到了这句话。最近基于注册和医院的研究记录了乳腺癌后黑色素瘤的风险在统计学上显著增加，标准化发病率从1.4到2.7不等。我在为非专业公众写关于这个话题的文章，我想验证这些数字，这样我就可以在我的故事中使用它们。因此，我从2009年的论文中提取了一些参考文献。当我提取他们的第五份参考文献，一篇2004年的论文时，我发现了与我在这里强调的完全相同的句子。2009年的作者逐字逐句地复制了一、二、三和四供参考。这些参考文献与2004年作者的摘要完全相同。现在，这很微妙，公平地说，我在报纸的其他地方没有发现任何其他抄袭的证据。但这是抄袭，它有两个问题。首先，2009年的作者借用了以前作者的确切措辞。以前的作者仔细阅读了文献，总结了所有内容，对其进行了综合并用文字表达。而且2009年的作者只是在偷那个。第二个问题是，2009年的作者完全相信2004年的作者已经准确地总结了文献。这是我不会接受的信心飞跃。正如我之前告诉你的那样，文献中普遍存在错误引文和不准确的引文。我永远不会相信这样的次要来源。我总是会回到主要来源，亲自检查数字。另外，2009年的作者假设在2004年至2009年之间没有发生任何新的事情。他们没有费心更新文献检索。现在，第二篇论文的作者可以直接将复制的句子放在引号中。这表明他们借鉴了先前作者的作品和文字。这也使他们在准确性方面摆脱了困境，因为通过引用，你是在说这些是别人所说的数字。你并不是说你已经核对或同意这些数字。但更好的是，2009年的作者应该做的是自己挖掘文献，从头总结文献，而不是简单地模仿别人的摘要。